NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

FORECASTING MARINE CORPS ENLISTED LOSSES

by

Sanford C. Orrick

March 2008

Thesis Advisor:	Samuel E. Buttrey
Associate Advisor:	Stephen L. Mehay

Approved for public release, distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2008	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE Forecasting Marine Corps Enlisted Losses

5. FUNDING NUMBERS

6. AUTHOR(S) Sanford Orrick

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

The Marine Corps has recently been authorized to increase end strength by approximately 20,000 Marines over the next 3 years. This has made forecasting of attrition an even more vital part of manpower planning. In order to successfully plan accessions to build the force we must be able to predict yearly attritions within the Marine Corps as accurately as possible. Because the enlisted force makes up the largest portion of the Marine Corps it is the most critical piece in accurately forecasting attritions.

This research compared end of active service (EAS) losses to non-EAS losses (excluding retirement). It used logit regressions to forecast losses with some success. It is not the final word in forecasting but rather a proof of concept in predicting such losses. All three of the models that were used to predict losses for fiscal years 2005, 2006, and 2007 had misclassification rates below 22 percent. This logit technique uses the attributes found in the models to predict a Marine's probability of becoming an NEAS loss. This logit technique does not take averages across years to predict losses; rather, it finds the attributes that are more likely to be associated with NEAS loss according to the data. This research is the beginning stage of what can ultimately be a model that looks at entry level recruits' attributes with an eye toward predicting if they will become NEAS losses in the future.

14. SUBJECT TERMS Forecasting, Attrition, Marine Corps NEAS losses

15. NUMBER OF PAGES 69

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

FORECASTING MARINE CORPS ENLISTED LOSSES

Sanford C. Orrick
Captain, United States Marine Corps
B.A., University of Missouri, 2001

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
March 2008

Author: Sanford C. Orrick

Approved by: Samuel E. Buttrey
Thesis Advisor

Stephen L. Mehay
Associate Advisor

Robert N. Beck
Dean, Graduate School of Business and Public Policy

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Marine Corps has recently been authorized to increase end strength by approximately 20,000 Marines over the next 3 years. This has made forecasting of attrition an even more vital part of manpower planning. In order to successfully plan accessions to build the force we must be able to predict yearly attritions within the Marine Corps as accurately as possible. Because the enlisted force makes up the largest portion of the Marine Corps it is the most critical piece in accurately forecasting attritions.

This research compared end of active service (EAS) losses to non-EAS losses (excluding retirement). It used logit regressions to forecast losses with some success. It is not the final word in forecasting but rather a proof of concept in predicting such losses. All three of the models that were used to predict losses for fiscal years 2005, 2006, and 2007 had misclassification rates below 22 percent. This logit technique uses the attributes found in the models to predict a Marine's probability of becoming an NEAS loss. This logit technique does not take averages across years to predict losses; rather, it finds the attributes that are more likely to be associated with NEAS loss according to the data. This research is the beginning stage of what can ultimately be a model that looks at entry level recruits' attributes with an eye toward predicting if they will become NEAS losses in the future.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGEMENTS

I would like to thank Professor Sam Buttrey, without your guidance this thesis would have never happened. Thanks for taking the time to explain an area of analysis, unfamiliar to myself, that made the entire thesis work.

I would also like to thank Professor Mehay for the continued guidance on this thesis and throughout my time at NPS.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

The Marine Corps has recently been authorized by Congress to increase its end strength by approximately 20,000 personnel over the next 3 years. This end strength increase over the next few years has made the forecasting of attrition a more vital part of manpower planning for the Marine Corps. In order to successfully plan its accessions to build the force, we must be able to predict annual attrition within the Marine Corps as accurately as possible. Because the enlisted force makes up the largest portion of the Marine Corps, it is critical that the forecasting models of enlisted attrition be as accurate as possible.

End strength is calculated at the end of each fiscal year as follows:

end strength = fiscal year beginning strength – losses + gains

End strength is mandated by Congress, and must not exceed 3% above the authorized end strength numbers. A 2% overage must be authorized by the Secretary of the Navy, and a 3% overage must be approved by the Secretary of Defense. This is the only tolerance allowed regarding end strength numbers. There is no authorized number for falling under end strength. However, if attrition is not accurately forecasted, it may lead to an underestimation of attrition, leading to insufficient new accessions, which in turn could bring operational consequences for the Marine Corps (Hattiangadi, Kimble, Lambert, and Quester, pp. 6-7).

The accurate forecasting of attrition has had an impact on the Marine Corps' annual budget. For instance, as of 2004, its progressively growing manpower cost was around $9.4 billion, about 60% of the Marine Corps' annual budget. If the Marine Corps does not accurately forecast attrition rates, it will have a cascading effect on the money spent on manpower, whether the forecast is over or under its annual budget. Because the budget is a constraint, it is very important that the Marine Corps' monthly forecasted attrition rates be as close as possible to the true numbers.

Overestimation of attrition rates leads to an unwarranted increase in accessions, thereby leading to an overspending on the Corps' annual manpower budget.

With non-end of active service (NEAS) losses accounting for 46 percent of all enlisted losses, and given the required increase in end strength over the next three years, it is very important to predict these losses as accurately as possible, as they will continue to have an effect on the Marine Corps' yearly accessions. The NEAS losses are broken down into three categories: (1) recruit losses, representing 12 percent of total losses, (2) retirement losses, representing 6 percent of total losses and (3) category losses, which represent the largest portion of losses at 28 percent. Each category is discussed below in the literature review section (Hattiangadi, Kimble, Lambert, and Quester, pp. 25-26).

1. Recruit Losses

Recruit losses are losses from Marine Corps Recruit Depot, San Diego or Parris Island. This category makes up 12 percent of the enlisted losses. Recruit losses are currently forecast by looking at the historical recruit loss rates for the previous four years. This is obtained by using the number of losses for each month, divided by the number of phased accessions for that month, to obtain a percentage for that particular month. The loss rates are averaged, then years are weighted - weighting of the years is determined by the planner - to get the predicted loss rates for the next fiscal year (Hattiangadi, Kimble, Lambert, and Quester, p. 27).

2. Retirement Losses

Retirement losses make up six percent of enlisted losses each year. The Marine Corps' retirement loss forecasting is done by capturing, in the month of September, all of the planned retirements for the previous fiscal year. Once this data is received, the planner removes all of the physical disability retirements, totaled in the categorical loss forecast, from that data. The remainder is now the base for the projection of the upcoming fiscal year. Because the planners are only getting the number of planned retirements from the previous fiscal year to use as a forecast for the following year the total number of forecasted retirements is usually low. To account for this, planners try to even out the shortage by calculating the average percentage of overage for the four

previous fiscal years. This average is calculated by comparing the number of planned retirements for each fiscal year to the actual retirements in that fiscal year. Once this average is calculated, it is then applied to the planned retirement number. This planned retirement number is then broken down into monthly retirement forecasts by looking at the historical averages by month for the previous four fiscal years. This historical monthly average is also looked at as a percentage of the previous four fiscal years' total number. It is this percentage by month that is applied to the planned retirement number to get the planned monthly retirement numbers (Hattiangadi, Kimble, Lambert, and Quester, pp. 32-35).

3. Category Losses

Category losses make up 28 percent of all of the enlisted losses each year. This categorical loss subsection is further divided into six sections within the category of losses. These are as follows: Convenience of the government, physical disability, misconduct, unsatisfactory performance, deserter status, and death (either combat or non-combat). The Marine Corps uses two methods to forecast the category of losses. One method is a steady-state model that predicts the monthly NEAS category losses using weighted averages. This type of forecasting is done with a steady inflow of yearly accessions and predicted losses. The second method is done using a Monte Carlo simulation. This simulation uses weighted averages as well. The value given to the weights can be adjusted by the manpower planner running the simulation. In many instances, the same values used in recruit losses weighted averages are used in the Monte Carlo simulation for category losses (Hattiangadi, Kimble, and Lambert, pp. 36-40).

The inaccurate forecasting of the Corps' NEAS losses could, again, lead to a miscalculated accession number that leads to overspending, if the forecasted NEAS losses are too high, or an undermanned goal, if the forecasted NEAS losses are too low.

B. PURPOSE

The purpose of this thesis is to examine the current methodology of forecasting enlisted loss rates in the Marine Corps. The Thesis also proposes to improve the ability to accurately forecast non-end of active service (NEAS) attrition. Given the required increase in end strength over the next three years, forecasting losses within the enlisted ranks will become an even more crucial aspect of manpower planning. As this research also entails an attempt to predict human behavior, forecasting such attrition rates is found to be a challenging task.

This research attempts to model, more accurately, the causal factors associated with the Marine Corps' enlisted ranks who depart the Marine Corps before their End of Active Service (EAS). The model is formulated to better forecast their NEAS attrition by researching and choosing attributes that may be significant predictors of the probability of attrition. The research done for this thesis focuses on the questions below.

1. Primary Research Questions

1. What factors and methods are currently used to predict enlisted non-EAS loss in the Marine Corps?

2. Can a model be developed that can help better predict enlisted non-EAS losses in the Marine Corps?

C. SCOPE AND METHODOLOGY

Although it is impossible to eliminate NEAS attrition, the Marine Corps would like to keep it at a minimum. Because NEAS attrition accounts for a large percentage of USMC enlisted losses each year, about 46%, the scope of this research will be to focus on this category of losses to better understand how to identify Marine Corps personnel that may fall into this category. The data used for this research was obtained from the Total Force Data Warehouse. It includes three different sets of data captured by fiscal year. The first is accession data from 1997 to 2007. The second data set used in this research is all end of active service and non-end of active service losses between 1997 and April 2007. This data set is broken down to compare Marines who left the service at their EAS

to the Marines who left the service before the end of their obligated service and are categorized as an NEAS loss. The third data set is an end-strength snapshot for fiscal year 1997.

This data provides empirical evidence of the attributes of someone who is likely to leave the service before the end of his or her current contract. This is accomplished by comparing the attributes of those Marines in the data set who complete their obligated service and are categorized as an EAS separation to the attributes of those Marines who do not complete their obligated service and are categorized as an NEAS loss. The analysis of the empirical data will identify the individual characteristics that predict a greater propensity of leaving the service early. In turn, it will be easier to forecast attrition behavior of those holding such characteristics in the future.

D. ORGANIZATION OF THE STUDY

Chapter II of the study is a literature review of the previous research done on attrition and a discussion about the current forecasting models used by Headquarters Marine Corps. Chapter III describes the data used to conduct this research. This chapter defines each variable used in the model, and gives the descriptive statistics for the data used in the logistic regression models. Chapter IV defines the logistic model and discusses the model's specifications in depth. Chapter V summarizes the results of the thesis and makes recommendations for further research in the area of forecasting the Marine Corps' NEAS losses.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. PREVIOUS ATTRITION AND LOSS STUDIES

The study of the Marine Corps' attrition and loss rates and its causes has been an ongoing theme since the inception of the all-volunteer Marine force in 1973. The attrition of first-term Marines has had a far-reaching effect not on only recruiting, but also on budgeting. The ability to accurately forecast attrition and losses is essential to minimizing the possible overspending of the budget on accession, as well as helping the recruiting force by getting the true numbers needed to recruit from month to month. It is also costly to the Marine Corps as an organization. There is no return on investment in man-hours spent training a first-term Marine if he or she departs before the end of his or her obligated service.

The Marine Corps, however, is concerned not only with first-term attrition rates; it must also account for those Marines who leave the service after their first term of service, whether at end of active service (EAS) or otherwise. This category of Marines is also accounted for when it comes to forecasting the next year's accessions, and if the number of losses is poorly predicted, there are implications for budgeting, as well as for the end-strength numbers.

Although there have been many studies that analyzed attrition, there are fewer studies that actually look at improving current methods of forecasting attrition, as well as the losses of those who leave at the end of their first-term of service. This chapter discusses four previous studies on the topic of attrition, as well as the ability to accurately forecast attrition. Included in these four studies is the report done by the Center for Naval Analysis (CNA) titled, "End-strength: Forecasting Marine Corps Losses Final Report" (Hattiangadi, Kimble, Lambert, & Quester, 2005).

This study looks at the Marines Corps' current procedures for predicting attrition, as well as losses. The Marine Corps currently uses weighted averages, moving weighted averages, and exponential smoothing in forecasting categorical losses. The CNA study looks at each category of loss and tries to enhance the current methods used by the

Marine Corps to get a more accurate prediction of future losses. The CNA study introduces a new method as one of their recommendations: the use of simple regression to forecast future losses. Although this method is recommended there is no use of regression models in the study.

The second study, a Naval Postgraduate School thesis, looks at first-term attrition rates among Marines by using survival analysis methods, as well as logit regression models. Survival analysis is used because of the nature of first-term attrition, compared to separation of those who complete their first term of obligated service. Once this analysis is run, the variables are then modeled using logit regression to look at the fit of predictions (Hawes, 1990).

The third study was chosen because it not only looked at a different population of Marines, but it also used the binary choice, or logit, model in an attempt to predict future attrition rates among that population. This study was also a Naval Postgraduate School thesis. In this study, the authors chose to use the logit model to forecast Marine officer attrition. The study breaks down the sample into six subcategories and models each separately. Although this is not the subject of my Thesis, it was chosen to get a look at the behavior of the logit model on a different population of Marines. Because so many studies have been done on predicting enlisted behavior, or the decision to attrite, it alos was chosen to get a better understanding of how well, or poorly, the model predicts when given a sample from a different population (Hurst & Manion, 1985).

The fourth study analyzed retention in the United States Marine Corps Reserves by using the logit model. This study was chosen to look at a different population and to see how the logit model's outcomes differ with this population when it attempts to forecast the decision to stay in or leave military service (Schumacher, 2005).

1. Hattiangadi, Kimble, Lambert, and Quester (2005)

This 2005 CNA report discussed the current methods used by Marine Corps Manpower Planners to forecast attrition, as well as losses. Attrition is defined as any time a Marine departs before his or her first term of obligated service is completed. The CNA report attempts to analyze the current procedures for forecasting each category of

either attrition or EAS losses, and provides recommendations on ways to possibly improve the Marines Corps' ability to forecast each. It is in no way a quick fix to the current forecasting situation, but it provides insight into the possible solutions that may help manpower planners forecast future attrition and losses. For the purpose of this thesis, the focus will be on the current NEAS procedures and the proposed recommendation made by the CNA report to use a regression model in an attempt to forecast attrition.

With NEAS losses currently accounting for approximately 46 percent of all enlisted losses, and given the increase in end-strength numbers over the next three years, it is very important to predict these losses as accurately as possible, as these losses will have an even greater effect on the Marine Corps' yearly accession goals in the future. NEAS losses are broken down into three categories: recruit losses, retirement losses, and category losses.

The forecasting of recruit losses is done by looking at the historical recruit loss rates for the previous four years. Because the recruit loss model assumes that each recruit is lost in the month in which he or she ships, it is recommended that there be a percentage assumed to be lost in the shipping month and the remaining percentage calculated as lost in future months (beyond the month shipped to boot camp). This method spreads unusually high loss numbers across the months.

The next recommendation made by the CNA study is that the exponential smoothing model be used, giving most recent data the heaviest weight, and progressively less weight to older observations. The problem with this is that there may not be the same behavior among the enlisted population from year to year, so the exponential smoothing model could still miss the mark when it comes to forecasting recruit attrition rates.

The CNA study also recommends the use of an optimization tool, such as the one used by the United States Air Force, to forecast those attrition rates. This optimization tool looks at previous years' known attrition numbers.

Those attrition numbers are then analyzed to determine exactly what weight to give each month based on the available historical data.

The recommendation for improving current retirement loss forecasting is to add unemployment rates to the current method used by the planners. The theory behind using the unemployment rates to predict a Marine's decision to stay in or leave (retire) is not a new one. It has been shown that when unemployment rates are low, there is a greater propensity for someone to leave the Corps and when unemployment rates are high, the propensity is reduced. The CNA study shows that adding the unemployment rate to the model produces predictions much closer to the actual retirement numbers than those from the method currently used by the manpower planners. Category losses make up 28 percent of all of the enlisted losses each year. This categorical loss subsection is further divided into six sections within the category of losses: convenience of the government, physical disability, misconduct, unsatisfactory performance, deserter status, and death, divided into either combat or non-combat death category losses.

The Marine Corps uses two methods to forecast category losses. One method is a steady state model that predicts the monthly NEAS category losses using weighted averages. This type of forecasting is done with a steady inflow of yearly accessions and predicted losses. The second method is done using a Monte Carlo simulation. This simulation uses weighted averages as well. The value given to the weights can be adjusted by the manpower planner running the simulation. In many instances, the same values used in recruit losses weighted averages are used in the Monte Carlo simulation for category losses.

The recommendation for improvement in the current methods of category losses takes into account the fact that end-strength numbers will grow over the next several years. The current method forecasts the number of categorical losses per month. Because the end-strength number is not going to be constant over the next several years, this could lead to a forecasted loss number that is much too low. The difference proposed by the CNA report is to forecast category losses by an average rate by month taken from the previous three years of known category losses.

Although not all of these recommendations by the CNA report may be implemented, they provide a possible starting point for enhancement to the current methods. No method used will produce 100 percent accuracy when it comes to forecasting attrition and losses. It is, however, a worthy goal to try and get as close as possible to the real numbers for the reasons stated earlier.

2. Hawes (1990)

A Naval Postgraduate School thesis, written by Eric Hawes in 1990, examined the attrition rates of first-term Marines by using survival analysis techniques. The use of survival analysis techniques is most prevalent in medical studies. It looks at failure times among the participants. As such, the use of survival analysis as a technique to determine attrition behavior among first-term Marines is not unlike the use of the survival analysis approach that is often used in medical studies.

A Marine's failure time is calculated as the amount of service completed prior to attrition. Marines who complete their first term of service, or fall into a special circumstance of early release from service, are handled as "censored" observations. The advantage in this type of attrition modeling is that all of the data, including the censored observations, can be used. It also allows censored observations to be separated from those that actually failed or attrited early.

The data used in this study was based on male, first-term recruits, with no prior service, who accessed between October 1, 1983 and September 1988. This collection was about 99 percent of accessions for that time period taking into account no female observations. The one drawback to this sample was that it did not include a representation of the female population of that same time period and an explanation to this missing data was not given. This is a flaw in the analysis of the thesis.

The data was broken down into three groups of covariates: (1) education credentials (Tier I being high school graduates, Tier II being alternate high school credential holders, and Tier III being non-high school graduates); (2) Armed Forces Mental Group (AFMG), (I, II, IIIA, IIIB, IVA, IVB, V), and (3) presence or non-presence of moral wavier. These were then analyzed in the thesis to see the effects of

each separately, as well as together. The author then broke down the sample into cohorts by fiscal year to perform the analysis for each year separately. This separation of cohorts into fiscal years could break out any attrition trends among the subgroups in the larger, pooled sample.

The results of this survival analysis are not unlike those of other attrition studies. For the education credentials, for example, it was found that Tier I were less likely to attrite than those in Tier II. The results further indicated that Tier III enlistees were more likely to attrite than the rest of the sample. It was also found in the single covariate results that there was a strong correlation between the mental group and the predicted probability of attrition. The higher the mental group, the less likely it was that the Marine would attrite. The moral wavier covariate was no surprise either; those with a moral wavier were more likely to attrite than those with no moral wavier.

In the survival analysis with combined covariates, holding education constant, Marines in Tier I and Tier II were more likely to survive (or not attrite) than those in Tier III, or than even those in Tier III with higher aptitudes. With the education level held constant, including those with or without moral waivers showed that recruits from Tier I or Tier II were, again, more likely to complete service. Marines in Tier III with a moral waiver were more likely to attrite than those in Tier II with no moral waiver. It was also found that those in Tier II holding a GED or correspondence school certificate had attrition rates related to the amount of actual "seat time" in school. Again, this should come as no surprise, as there have been many studies done showing the correlation between the lengths of time actually spent in school being negatively correlated to the person's likelihood of leaving his or her service in the Marine Corps early (Hawes, pp. 18-44).

3. Hurst and Manion (1985)

In a Naval Postgraduate School thesis completed in 1985, Stephen Hurst and Thomas Manion studied the attrition rates of Marine Corps officers by using a binary choice, or logit, model. Their thesis builds off of a previous thesis that looked at officer attrition, and included in the model economic factors, more specifically, unemployment

rates and pay grades, as well as promotion potential based on fitness report data. The authors used a logistic model to predict officer attrition (Hurst & Manion, p 9).

The authors used Military Occupational Specialty (MOS) (ground or air), pay grades, Armed Forces Active Duty Base Date (AFADBD), type of degree (either technical or non-technical), and a self-developed variable built from fitness report data, called performance index score. The performance index score is a variable built to quantify the officer's promotion potential based on fitness report data input. This is a very subjective variable, but it is needed as a proxy for performance (Hurst & Manion, pp 9-13).

The models were divided into military rank, and then each rank was divided into either ground or air MOS's to eliminate as much variability as possible in the model. One exception to this was the rank of Colonel which was looked as an entire sample with no division into ground or air. This was done because once a Marine is promoted to the rank of Colonel, his or her MOS no longer distinguishes between ground and air.

The data collected for this study was from the Manpower Management System. It consisted of 132,903 records of officers from fiscal year 1977 to fiscal year 1984. The authors tested their binary choice's model predictions against the actual attrition rates for fiscal years 1981 and 1982. The explanation for this was that it was easier to show the results of comparison in their study for two specific years rather than all eight. Although this approach eases the burden of work for the study, this approach may not show a true representation of the attrition behavior of the Marine Corps officers.

The study showed that younger officers were less likely to leave the Marine Corps when unemployment rates were high. This is not surprising, as the younger officer may see the Marine Corps as a secure employment opportunity and, therefore, not leave the service. It was also found that those with the higher performance index rating (based on fitness report data) were more likely to stay in, and those with the lower score were more likely to leave. This may be due to the fact that this performance rating score is tied to promotion potential and, therefore, those with lower scores are not being selected for promotion beyond the rank of Captain. This would lead to the Marine being forced out

of the Marine Corps. The binary choice model forecasted 186 lost officers for fiscal year 1981, compared to the actual attrition of 183. The forecast for 1982 fell short, with 140 losses being forecast, compared to the actual attrition of 169.

The sample of Captains showed a more significant difference when it comes to pay and college major. For those who were in ground MOS's, a higher performance index led to a greater probability of staying in the Marine Corps. This is again, tied to promotion potential. This model also showed that it was more likely that those holding technical degrees (engineering for example) would leave than those with liberal arts degrees. Captains in the aviation community put a bigger emphasis on pay as the deciding factor to stay. The education variable was not used for aviation Captains because of the inherently technical nature of piloting. The significance of the pay may be an underlying factor in the decision to pay bonuses to pilots who could leave the service for a much more lucrative civilian career as a pilot. The model predicted 50 losses for 1981, with the actual being 50. The model predicted 32 losses for 1982, compared to 43 actual losses.

In the sample of Lieutenant Colonels, it was found that for the ground community, MOS's within that community was a deciding factor. It is a key point that receiving command time at every level is perceived to be an important qualification, and there is little command opportunity for those in more restrictive ground MOS's. This could lead to the officer seeing his or her chances of promotion to a higher level being smaller than those with command time.

The variables representing pay, education major, and unemployment rate became insignificant at this level. Presumably if pay were an issue, the officers in question would have left many years ago, so pay would not have been a factor in their decision to leave. The curious finding at this level is that the performance index rate at this level had the opposite effect on retention. Those with higher performance index rates were leaving the Marine Corps at a higher rate. The model predicted 45 losses for Lieutenant Colonels in ground MOSs, with the actual loss of 34. The model predicted 14 losses for Lieutenant Colonels in aviation MOSs, with the actual loss of 13. Again, the performance index rate had the opposite effect for those in aviation. For the Colonels sample, the model showed,

again, that those with higher performance ratings were more likely to leave the Marine Corps. This sample, however, may be harder to predict, no matter the results, as these officers were beyond the 20-year mark and retirement eligible. Therefore, many more factors could go into the decision to stay in or leave the military service.

In summary, the prediction rates for this study were not encouraging. Even though losses forecasted by the 1987 models were within 90 percent in 3 of 5 cases, accuracy fell off drastically in the fiscal year 1982 models. In that case the accuracy of all five models was far below the 90 percent level. There is no explanation as to why the fiscal year 1982 predictors missed the mark so badly, and the 1981 predictors are all very close to, if not the same, as actual. The reality is that in the officer community, there is almost a set career track that must be followed. If this career path is not followed, promotion becomes less likely. Any perceived notion by an officer that his or her promotion potential is small may lead to a decision to leave the Marine Corps.

This 1985 study still, however, provided a good foundation on how the binary choice model worked when applied to a sample of those who made the choice to leave their service in the Marine Corps. Even though there was no explanation given for the differences in each of the predictions (when compared to the actual numbers) the binary choice model's outcome may have a lot to do with the makeup of the sample itself.

4. Schumacher (2005)

Because of the recent increase in operational tempo over the past five years as a result of Operation Enduring Freedom and Operation Iraqi Freedom, there has been a greater need to call the reserves force to active duty. This includes entire units, as well as individual reservists to fill "individual augmentation" billets overseas. In this Naval Postgraduate School thesis, the model showed the impact of mobilization and unemployment on an individual's decision to stay in or leave the Marine Corps Reserves. The goal was to better establish recruiting and retention goals for the reserves population.

Bureau of Labor and Statistics (BLS) and Reserve Component Personnel Data are used, as well as mobilization data from Defense Manpower Data Center (DMDC). The author hypothesizes that there is a correlation between the number and length of

activations and the decision made to stay in or leave the reserves. The base individual used by Schumacher is a single male, with no dependents, no mobilization time, and zero years of active service.

In determining the likelihood of whether a Marine will stay in or leave the reserves Schumacher used a model that included sex, number of dependents, years in service, length of time mobilized, number of mobilizations, months served in a reserve category, and yearly home of record state unadjusted unemployment rate at the end of service. Each of the variables in the model was found to be significant at the .01 level, with unemployment rate and number of months mobilized having a negative effect on the recruit's decision to stay in the reserves, as each of the two variables increased.

This study finds that short call-ups for reserve Marines has a positive effect on his or her decision to stay in the reserves force. However, the opposite is true when it comes to longer active tours of duty. Among those who are called to active duty for longer periods of time, it is more likely they will leave the reserves.

There were variables missing from this study that have been shown in previous studies to impact retention behavior. In particular, the omitted variables included rank, marital status, and the educational level of the individual. These three variables have proven, in previous studies, to show some explanatory value when it comes to the decision to stay in the service, either active or reserves. Thus, this may limit the findings of the study.

Although this study had its limitations in the number of explanatory variables used, it is no surprise that the time spent mobilized had a negative effect on the Marine's decision to stay in the reserves. If the Marine wanted to be on active duty for a longer period of time, he or she would have joined the active force, and not the reserves. The other explanatory variable that had a negative effect on the decision to stay in the reserves is the unemployment rate at the end of service. This, too, is of no surprise, as it followed the behavior of many attrition studies done on the active force (Schumacher, 2005.)

III. DATA AND METHODOLOGY

A. INTRODUCTION

This chapter discusses the data used in the statistical analysis of non-end of active service losses and attrition in the Marine Corps. It discusses the data collection process and gives a short summary of the data collected together with descriptive statistics. The methodology used to forecast non-end of active service losses is also discussed. The analysis of the data collected will help identify attributes that may lead to a better forecast of attrition or losses within not only the first-term population but the population of careerists as well.

B. DATA COLLECTION

The data in this study is from the Marine Corps' Total Force Data Warehouse (TFDW). The collection itself consisted of three different sets of data. The first data set captured all enlisted losses from the period of October 1, 1997 to April 30, 2007. The second data set captured all enlisted accessions from the period of October 1, 1997 to April 30, 2007. The final data set provided a snapshot of enlisted end strength ending on September 30, 1997. The end strength data is used to capture attributes for those enlisted losses that may not be captured in the accessions data. Those three data sets were then merged into a single file for the statistical analysis.

C. DATA SUMMARY

The master data file compiled from the three merged data sets (losses, accessions, and end strength) was converted from the Microsoft Excel format into the DTA format for use in the STATA program for coding, cleaning and analysis. Entries that could not be relied on as accurate information were deleted. The merged file consisted of 587,154 entries. However, once the merged file was cleaned for inaccurate entries the final data set included 167,269 observations. The large difference is due to many observations being omitted for reasons such as missing separation codes, or erroneous entries. This data does include observations missing variables such as race. Observations missing

variables and retained in the data set were given codes of “other” for their missing values. This data set was further divided into fiscal year based on each Marine’s end of active service date or end of current contract date. The creation of binary variables was done for logistic modeling. The data descriptions in Table 3.1 below shows the variables created from the data file and used to estimate the logistic regression models. All were generated from original data fields. This set of variables was further divided into fiscal years to compare differences across years. The separation categories were combined into all NEAS losses which were used as the binary dependent variable. The remaining variables represent binary independent variables.

Table 3.1. Data Description

Variables	**Description**
afqt5	=1 if Category V; 0 otherwise
afqt4	=1 if Category IV; 0 otherwise
afqt3b	=1 if Category IIIb; 0 otherwise
afqt3a	=1 if Category IIIa; 0 otherwise
afqt2	=1 if Category II; 0 otherwise
afqt1	=1 if Category I; 0 otherwise
male	=1 if missing race category; 0 otherwise
onedependent	=1 if one dep; 0 otherwise
two dependents	=1 if two dep; 0 otherwise
three dependents	=1 if three dep; 0 otherwise
four dependents	=1 if four dep; 0 otherwise
five dependents	=1 if five dep; 0 otherwise
six_more dependents	=1 if six or more; 0 otherwise
no dependents	=1 if no dep; 0 otherwise
no dependent information	=1 if missing dependent information; 0 otherwise
years_0_4	=1 if up to 4 years of service; 0 otherwise
years_4_8	=1 if 4 to 8 years of service; 0 otherwise
years_8_12	=1 if 8 to 12 years of service; 0 otherwise
years_12_or_more	=1 if greater than 12 years of service; 0 otherwise
age_17	=1 if 17 years of age at enlistment; 0 otherwise
age_18_19	=1 if 18 to 19 years of age at enlistment; 0 otherwise
age_20	=1 if 20 years of age or older at enlistment; 0 otherwise
reenlist_retired	=1 if retirement; 0 otherwise
mcd_1	=1 if accession from 1st Marine Corps District; 0 otherwise
mcd_4	=1 if accession from 4th Marine Corps District; 0 otherwise
mcd_6	=1 if accession from 6th Marine Corps District; 0 otherwise
mcd_8	=1 if accession from 8th Marine Corps District; 0 otherwise
mcd_9	=1 if accession from 9th Marine Corps District; 0 otherwise
mcd_12	=1 if accession from 12th Marine Corps District; 0 otherwise
jr high_educ	=1 if junior high school education; 0 otherwise
high school_educ	=1 if high school education; 0 otherwise
college_educ	=1 if college level education; 0 otherwise
master_educ	=1 if masters degree obtained; 0 otherwise
postmaster_degree	=1 if postmasters degree obtained; 0 otherwise
doctorate_degree	=1 if doctorate degree obtained; 0 otherwise
legal_separated	=1 if legaly separated; 0 otherwise
not_married	=1 if not married; 0 otherwise
married_other	=1 if marital status not reported; 0 otherwise
retirement separation	=1 if retirement sep; 0 otherwise

Table 3.1 continued

Variables	Description
unsat performance separation	=1 if unsat sep; 0 otherwise
deserter separation	=1 if deserter sep; 0 otherwise
physical disability separation	=1 if phy disab sep; 0 otherwise
court martial	=1 if court martial sep; 0 otherwise
enlisted to officer separation	=1 if enl to off sep; 0 otherwise
misconduct separation	=1 if misconduct sep; 0 otherwise
con of govt separation(cov)	=1 if COV sep; 0 otherwise
eas separation	=1 if EAS sep; 0 otherwise
contract_4yr	=1 if 4-year contract signed; 0 otherwise
contract_5yr	=1 if 5-year contract signed; 0 otherwise
contract_6yr	=1 if 6-year contract signed; 0 otherwise
contract_3yr	=1 if 3-year contract signed; 0 otherwise
contract_8yr	=1 if 8-year contract signed; 0 otherwise
adult diploma	=1 if adult diploma obtained; 0otherwise
occupational certificate	=1 if occupational cert completed; 0 otherwise
hs diploma	=1 if diploma obtained; 0 otherwise
less high school	=1 finished less than high school; 0 otherwise
ged	=1 if GED completed; 0 otherwise
home school	=1 if home school complete; 0 otherwise
college degree	=1 if college degree complete; 0 otherwise
one semester college	=1 if one semster complete; 0 otherwise
high school senior	=1 if not graduated; 0 otherwise
other_school	=1 if missing category; 0 otherwise
no combat tour	=1 if missing category; 0 otherwise
combat_tour	=1 if completed combat tour; 0 otherwise
american indian	=1 if American Indian; 0 otherwise
asian_pacific islander	=1 if Asian or Pacific Islander; 0 otherwise
otherrace	=1 if missing race category; 0 otherwise

Source: created by author from data

D. DESCRIPTIVE STATISTICS

Descriptive statistics for all variables are shown in Table 3.2. The distribution of AFQT scores matches normal USMC recruiting patterns. The gender variable shows over 90 percent of recruits are male. The dependents variable is included in the data although 55 percent of the observations are missing this information. The decision to

leave it in the models rested on the idea that there is enough variation among those that do have this information to perhaps show significance in the models.

The years of service variable is divided into four categories: those serving up to 4 years, those serving between 4 and 8 years, those serving between 8 and 12 years, and those serving greater than 12 years. This is included to get a sense of time served at loss. The majority of enlistees served between 0 and 4 years. This category represents over 62 percent of the analyzed data. The smallest portion, at just over 4 percent, was those serving between 8 and 12 years.

The age at enlistment variable again was created to analyze the age of those being lost. The proposed effect is those that are younger at enlistment are more likely to become an NEAS loss. Over 68 percent of the observations are between ages 18 and 19 at enlistment.

The districts variable distribution is fairly uniform for five of the six USMC recruiting districts. The sixth, the 4th Marine Corps district, with 24,400 observations, was lower than the others. The variable for missing district information was labeled "other" and numbered only 1,091 observations, less than one percent. The overall distribution of this variable is as expected as each district is responsible for a roughly equal number of accessions each year.

The marital status and contract length variables are in line with the normal population of recruited Marines. Both observations seem representative of the average population of Marines recruited into the Marine Corps. A majority, 59.89 percent, of the sample was single. Over 80 percent of the observations signed a four-year contract.

The combat tour variable is another that has over 51 percent of its observations missing. It was retained in the models to see if there was a detectable difference among the 22 percent that did report serving or not serving in combat. This combat tour variable represents not only Operation Enduring Freedom and Operation Iraqi Freedom but includes any operation classified, by the Marine Corps, as combat that a Marine may have been involved in during his enlistment.

The education code and education certificate are variables that seem inconsistent. It appears that some of these variables are used interchangeably. There are 158,910 observations whose education code specifies a high-school education with another 58,173 reported as holding a high school diploma in the education certificate variable. Over 8,245 education codes report some form of college education, but it is reported in the education certificate that 5,173 observations have at least one semester of college or more. Because of this ambiguity, each code is used in the models. While this may create some overlap, it ensures there is no education category missed.

The separation codes are broken down according to the Marine Corps separation code definitions. Over 60 percent of the observations were reported in the EAS separation variable as having served honorably. This is a higher percentage than stated for EAS separations in Chapter 2. The difference in my research may be due to the amount of observation deleted because of missing separation code. The retirement separation code describes those Marines who retired after 20 years or more of service. The "convenience of the government" separation code represents 5 percent of the observations and includes sole survivors, hardship discharges, and conscientious objectors. The "misconduct" separation code represents 7 percent of the observations and includes those with drug offenses, minor disciplinary infractions, and patterns of misconduct. The "unsatisfactory performance" separation code represents 0.5 percent of the observations and includes weight control, unsatisfactory performance, unsanitary habits, and unsuitability. With a total of 10 percent of observations the recruit separation variable represents a majority of the NEAS losses in this study. The remaining separation codes are explained by their title in the table.

It must be noted that over 126,000 observations were missing a separation code. This amount of missing observations may have an influence on the outcome of the models. The separation code assigned at release from active duty has been shown to be very unreliable. This is due to the nature of reporting these codes. It is many times the administration clerk's responsibility to assign such a code and he or she may not be a reliable source of this information. However, no other source of information for this data is available for this study.

The final code is the race variable as reported by the Marine Corps. This variable was particularly difficult to interpret because the Marine Corps has changed its coding in recent fiscal years. Each of the letters represents different races depending on the fiscal year in which they were recorded. There were over 16,000 observations that denoted failure to respond or were missing a race code. Although this variable is missing many observations it was kept in the data set in lieu of the ethnicity code which was missing in more than half of the observations.

Table 3.2. Observations and percentage	Frequency	Percentage*
AFQT Scores		
afqt5	11	0.01
afqt4	1,382	0.83
afqt3b	52,272	31.25
afqt3a	44,211	26.43
afqt2	57,082	34.13
afqt1	6,026	3.60
no afqt score	6,285	3.76
Gender		
male	156,091	93.32
female	11,178	6.68
Dependents		
one dependent	3,532	2.22
two dependents	21,728	13.66
three dependents	4,053	2.55
four dependents	16,552	10.41
five dependents	1,077	0.68
six_more dependents	2,169	1.36
no dependents	24,492	15.40
no dependent information	93,666	55.99
Years of Service		
years_0_4	104,801	62.65
years_4_8	34,322	20.52
years_8_12	7,386	4.42
years_12_or_more	20,759	12.41
Age at enlistment		
age_17	8,270	4.94
age_18_19	114,525	68.47
age_20	44,474	26.59
District		
mcd_1	28,161	16.84
mcd_4	24,400	14.59
mcd_6	28,429	17.00
mcd_8	26,885	16.07
mcd_9	29,960	17.91
mcd_12	28,343	16.94
mcd_other	1,091	0.65
Education Code		
grade school_educ	7	0.00
midschool_educ	2	0.00

Table 3.2 continued	Frequency	Percentage*
jrhigh_educ	47	0.03
highschool_educ	158,910	95.00
college_educ	8,113	4.85
master_educ	112	0.07
postmaster_degree	15	0.01
doctorate_degree	5	0.00
Marital Status		
married	36,378	21.75
legal_separated	43	0.03
not_married	100,184	59.89
married_other	30,664	18.33
Separation Code		
retirement separation	8,992	5.38
unsat performance separation	943	0.56
deserter separation	4,415	2.64
recruit separation	16,857	10.08
physical disability separation	8,013	4.79
court martial	160	0.10
enlisted to officer separation	2,375	1.42
misconduct separation	12,237	7.32
con of govt separation(cov)	9,135	5.46
eas separation	104,142	62.26
Contract Length		
contract_4yr	167,269	81.48
contract_5yr	19,366	11.58
contract_6yr	4,500	2.69
contract 3yr	7,098	4.24
contract_8yr	1	0.00
Education Certificate		
adult diploma	1,749	1.05
occupational certificate	318	0.19
hs diploma	58,173	34.78
less high school	66	0.04
ged	3,847	2.30
home school	257	0.15
college degree	4,392	2.63
one semester college	781	0.47
other_school	1,375	0.82
Combat Tour		
no combat tour	44,291	26.48
combat_tour	36,277	21.69

Table 3.2 continued	Frequency	Percentage*
missing combat tour data	86,701	51.83
Race		
american indian	1,738	1.04
asian_pacific islander	2,855	1.71
african american	22,512	13.46
caucasian	123,723	73.97
hispanic	348	0.21
otherrace	16,093	9.62

N=167,269
*percents may not add to 100 because of rounding error
Table created by author from TFDW data

E. METHODOLOGY

Because of the binary nature of the attrition outcome the logistic regression model is chosen to forecast NEAS loss versus EAS separation. The logistic models are created using the binary dependent variable denoting the Marine's loss or attrition code. This dependent variable was then compared to a number of independent variables, chosen by the author, to try and identify attributes that distinguish between NEAS loss and EAS separation. For the purpose of this study the loss categories of death, whether accidental or combat related, and retirements were dropped from the sample.

The estimated model is specified as:

$$\mathrm{Ln}(p/1 - p) = \beta_0 + \beta_1 x_1 + \ldots \beta_k x_k$$

where p is the predicted probability that a Marine is an NEAS loss and 1 minus p is the predicted probability of being an EAS separation, β_0 is the intercept and β_1 through β_k are the predicted changes in the likelihood of becoming an NEAS loss given the independent variables, x_1 through x_k.

The model's independent variables are run against NEAS loss across all fiscal years. The model is then re-run using only data from fiscal years 1998 through fiscal year 2004. This model is used to get predictions of 2005 NEAS losses. The data is

widened by adding fiscal year 2005 to predict 2006 NEAS losses and again fiscal year 2006 data is added to get NEAS loss predictions for 2007.

Each of the three individual model's estat (EAS/NEAS) classifications were then run to see the overall correct classification. This post-estimation table compares the two types of losses present in the model and summarizes the statistics of all observations in the data showing the correct and incorrect classification of those observations. Each observation with a predicted probability greater than or equal to 0.5 is classified as an NEAS loss. Observations below .5 are classified as an EAS separation. The 0.5 classification threshold can be adjusted but was not done so for this research.

These classifications are shown in the estat classification tables which indicate the number of observations correctly identified as true NEAS losses (at or above 0.5 predicted probability and categorized as a NEAS loss by separation code) as well as those observations that are falsely identified as not an NEAS (below the 0.5 predicted probability but still categorized as an NEAS loss by separation code). The same calculations are done for EAS separations.

For each of the three logit models there was a Receiver Operating Characteristic (ROC) curve generated to see the overall performance of the models. A classifier whose ROC curve follows a 45-degree line has the same probability of classifying a positive observation as a positive as it does with a negative one. The ROC curve plots Sensitivity (probability of detecting true positives or NEAS losses) against 1 minus Specificity (probability of detecting true negatives or EAS separations), every possible value of the cutoff. As a general rule of thumb when the area under the curve (AUC) exceeds 0.8 the model is successful. The AUC can also be interpreted in this way: if one NEAS loss and one EAS loss are randomly chosen, the AUC gives the chance that the predicted probability of NEAS for the first observation exceeds that of the second.

Because the results for each of the three predicted years, 2005, 2006, and 2007 were so close in correct classification, Chapter IV only shows the results of this methodology for the data set containing fiscal years 1998-2004 to get the predictive probability of NEAS losses for 2005.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. MODEL ESTIMATIONS

A. MODEL

The logit model used to forecast the probability of NEAS loss where *p* is predicted probability of NEAS loss and 1 minus *p* is the predicted probability of an EAS loss is as follows:

$$Ln(p/1 \text{ minus } p) = \beta_0 + \beta_1(afqt5) + \beta_2(afqt4) + \beta_3(afqt3a) + \beta_4(afqt2) + \beta_5(afqt1) + \beta_6(nodep) + \beta_7(onedep) + \beta_8(twodep) + \beta_9(threedep) + \beta_{10}(fourdep) + \beta_{11}(fivedep) + \beta_{12}(six\_moredep) + \beta_{13}(female) + \beta_{14}(mcd\_1) + \beta_{15}(mcd\_4) + \beta_{16}(mcd\_6)\ \beta_{17}(mcd\_8) + \beta_{18}(mcd\_9) + \beta_{19}(mcd\_12) + \beta_{20}(gradeschool\_educ) + \beta_{21}(midschool\_educ) + \beta_{22}(jrhigh\_educ) + \beta_{23}(college\_educ) + \beta_{24}(master\_educ) + \beta_{25}(postmaster\_educ) + \beta_{26}(doctorate\_educ) + \beta_{27}(married)\ \beta_{28}(legal\_separated) + \beta_{29}(married\_other) + \beta_{30}(contract\_5yr) + \beta_{31}(contract\_6yr) + \beta_{32}(contract\_3yr) + \beta_{33}(contract\_8yr) + \beta_{34}(adult\_diploma) + \beta_{35}(occup\_cert) + \beta_{36}(less\_highsch) + \beta_{36}(ged) + \beta_{37}(home\_sch) + \beta_{38}(college\_degree) + \beta_{39}(sem\_college) + \beta_{40}(other\_school) + \beta_{41}(combat\_tour) + \beta_{42}(Amerindian) + \beta_{43}(asian\_pacislndr) + \beta_{44}(africanamerican) + \beta_{45}(otherrace) + \beta_{46}(Hispanic) + \beta_{47}(years\_0\_4) + \beta_{48}(years\_4\_8) + \beta_{49}(years\_8\_12) + \beta_{50}(age\_18\_19) + \beta_{51}(age\_20)$$

Although all fiscal year data was included in the analysis, the model above was restricted to fiscal years 1998 to 2004 to get a predicted probability of NEAS losses for the fiscal year 2005. The results of the logit regression are used as the foundation for tabulating the predicted probability of NEAS losses in FY2005.

B. LOGIT MODEL RESULTS FOR FY1998-FY2004

As seen in Table 4.1 a majority of the variables in the model are found to be statistically significant at the 1 percent level. The variable jrhigh_educ was found to be significant at the 5 percent level. The variables found to have no statistical significance are: afqt5, master_educ, legal_separated, less_highsch, sem_college, amerindian, africanamerican, hispanic. This may be due to the small number of observations for each of these variables in the data set.

Table 4.1. Coefficients of logit model

Non end of active service loss FY98-FY04	Coefficients	Z-stat
afqt5	2.146	-1.62
afqt4	0.505	(5.553)***
afqt3a	-0.129	(6.63)***
afqt2	-0.298	(15.51)***
afqt1	-0.421	(9.43)***
no dependents	-0.829	(38.11)***
one dependent	0.633	(13.52)***
two dependents	-1.764	(56.76)***
three dependents	-1.16	(25.81)***
four dependents	-3.075	(52.40)***
five dependents	-2.608	(18.61)***
six or more dependents	-1.507	(11.92)***
female	0.238	(8.20)***
mcd_1	0.355	(3.84)***
mcd_4	0.486	(5.25)***
mcd_6	0.441	(4.77)***
mcd_8	0.335	(3.62)***
mcd_9	0.316	(3.42)***
mcd_12	0.243	(2.63)***
jrhigh_educ	1.928	(1.80)*
college_educ	-0.427	(9.09)***
master_educ	-0.364	-1
postmaster_educ	0.242	-0.2
married	-0.517	(22.15)***
legal_separated	-0.373	-0.7
married_other	0.476	(26.52)***
contract_5yr	1.382	(44.65)***
contract_6yr	0.604	(11.02)***
contract_3yr	0.553	(9.16)***
adult_diploma	0.374	(5.25)***
occup_cert	1.008	(3.91)***
less_highsch	0.461	-1.55
ged	0.799	(16.40)***
home_sch	1.399	(5.30)***
college_degree	0.299	(4.56)***
sem_college	-0.138	-1.48

Table 4.1 continued	Coefficients	Z-stat
otherschool	0.348	(4.47)***
combat_tour	-1.057	(29.64)***
amerindian	0.08	-1.07
asian_pacislndr	-0.371	(5.93)***
africanamerican	-0.006	-0.26
other race	-0.331	(12.04)***
hispanic	-0.145	-0.6
up to 4 years of service	-3.263	(66.36)***
4 to 8 years of service	-4.926	(93.18)***
8 to 12 years of service	-3.857	(67.42)***
18 to 19 at enlistment	2.433	(2.82)***
20 or older at enlistment	2.759	(3.19)***
Constant	0.478	-0.55
Observations		105001
Absolute value of z-statistics in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%		

1. Estat Classification for FY1998-FY2004

Estat classification is a post-estimation STATA command run after the logit model. This function gives the correct classifications of NEAS loss compared to EAS separation as run by the model. The estat classification as seen in Table 4.2 represents the number of observations correctly identified as true NEAS losses (above .5 predicted probability and categorized as a NEAS loss by separation code) and then those that are falsely identified as not an NEAS (below the .5 predicted probability but still categorized as an NEAS loss by separation code). In this table the letter D represents NEAS loss and ~D represents EAS separation.

The results show that the NEAS loss was correctly classified 76.18 percent of the time. This can be compared to a correct classification of 62.41 percent using the naïve rate. The naïve rate is the overall sample rate of EAS separations, and is computed by adding the number of correctly classified EAS separation (9668) and falsely classified as EAS separations (55871), then dividing by the total number of observations (65539).

Comparing the naïve rate to the estat classification shows the model has generated an increase of 13.77 percent correct predictions.

Table 4.2. Estat classification FY1998-FY2004

Logistic model for neas_loss FY1998-FY2004			
	-------- True --------		
Classified	NEAS Loss	EAS separation	Total
+	24117	9668	33785
-	15345	55871	71216
Total	39462	65539	105001
	Classified + if predicted Pr(D) >= .5		
	True D defined as neas_loss != 0		
Sensitivity	Pr(+ D)	61.11%	
Specificity	Pr(-~D)	85.25%	
Positive predictive value	Pr(D +)	71.38%	
Negative predictive value	Pr(~D -)	78.45%	
False + rate for true ~D	Pr(+~D)	14.75%	
False - rate for true D	Pr(- D)	38.89%	
False + rate for classified +	Pr(~D +)	28.62%	
False - rate for classified -	Pr(D -)	21.55%	
	Correctly classified		76.18%

Output generated by STATA 9.1; table created by author

2. ROC Curve for FY05 Predictions

The ROC curve for Fiscal year 1998-2004 is shown in Figure 4.1. This ROC curve is generated with the assumption that every observation in the model with a predicted probability greater than or equal to 0.5 is an NEAS loss. This shows the model's overall ability to classify those that are NEAS losses against those that are

separated by EAS. The area under the curve for this model is .8591. The ROC curve shows that the model's assignment of probabilities is close to their actual value.

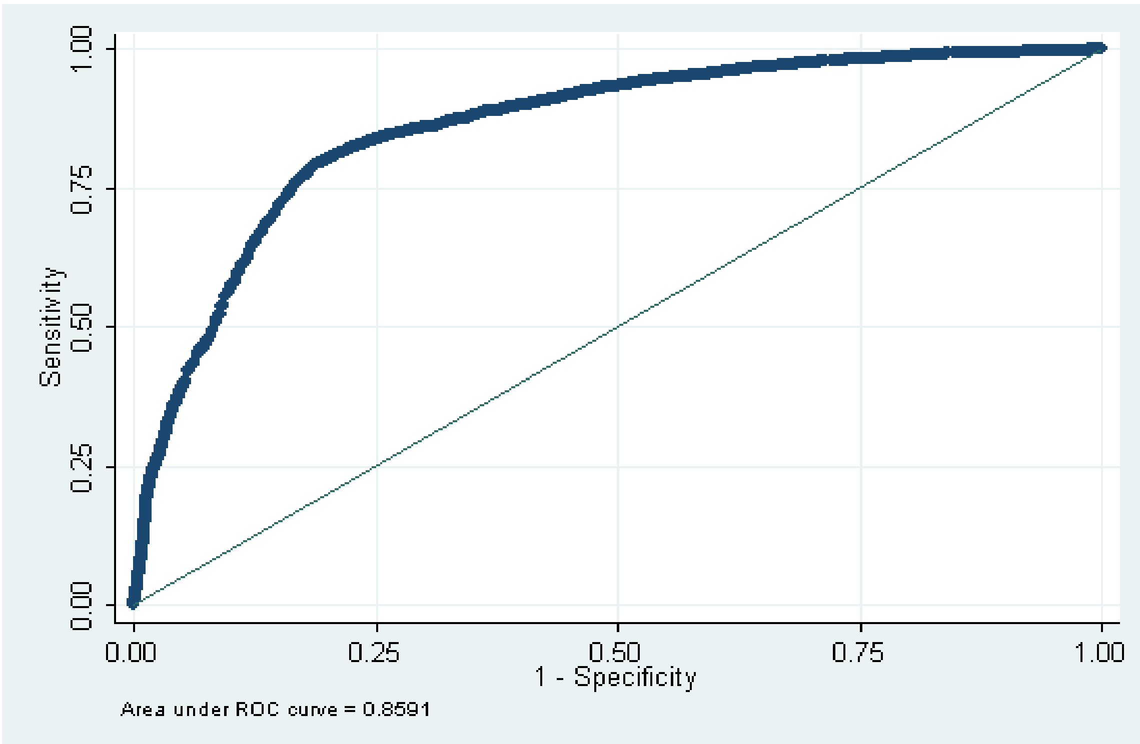

Figure 4.1. ROC curve for FY2005 predictions

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY AND RECOMMENDATIONS

A. SUMMARY

This Thesis developed a logit model to forecast NEAS losses of enlisted Marines by comparing NEAS losses to EAS separations. The logit model contained data that was broken down into each fiscal year according to the Marines' end of current contract date. The data included independent variables thought by the author (and based on the literature review) to be predictive of NEAS losses. Those independent variables were run against the dependent variable, NEAS loss, to predict the following year's loss rates. This model does not include any Marines who were still in the Marine Corps at the time of the research. The research only predicts EAS separations versus NEAS losses.

This logit model technique is an attempt at predicting losses using a method different from the one currently employed. It predicts loss types for a particular year based on attributes of the Marines leaving in that year. All three of the models correctly classified NEAS losses with greater than 76 percent accuracy and misclassified those that were EAS separations as NEAS losses at a rate below 25 percent. Receiving Operator Characteristics (ROC) curves show that the logit models perform well. Currently the Marine Corps does not use this type of forecasting for NEAS losses and before this forecasting method can be implemented further study must be done.

B. RECOMMENDATIONS

1. Forecasting by Separation Category

The models estimated in this research use NEAS losses, including both recruit losses and category losses. There may be some value in breaking down the NEAS loss variable into each of the separate losses found within the NEAS loss variable. The biggest proportion of this is the category loss. If the models can predict separation code based on attributes included in the data, more attention can be paid to those areas of separation. This may bring benefits in the future not only to manpower planners but also the Marine Corps as an organization.

The ability to identify those separations more likely to occur may help focus the efforts of trying to eliminate or lessen the propensity of a Marine separating for those reasons.

2. Forecasting by Military Occupational Specialty

Although there was an initial attempt in this study to include the effects of the MOS variable, obtained from TFDW, on losses the MOS data was largely missing. If accurate MOS data were available, a model run with the MOS variable included might well have improved performance. Such a model could also be developed into a standalone model that helps shape the population of Marines by MOS.

3. Forecasting by Month

An attempt to forecast losses by month should be made. This can be done with data that is broken down into each month of the fiscal year. This method of monthly forecasting can provide two things. First, it will allow the user to see differences in months not only within a fiscal year but among the fiscal years included in the model. Secondly, it will allow the user to see if there are any months more likely to have losses, and if this difference is constant across fiscal years. This monthly breakdown may help identify any seasonal influences. Once this is done steps can be taken to counter that seasonal influence.

4. Survival Analysis

The use of survival analysis was not attempted as part of this research. In an attempt to more accurately forecast NEAS losses survival analysis may be considered as part of future research. This technique has proven to be a very useful tool in its predictions based on attributes of a representative sample of the entire population. In the present case, data limitations would not allow this type of analysis. With the use of survival analysis the study can compare those Marines who are lost to those who survive throughout the study. This may help reach the ultimate goal of being able to develop a model that look at a population that has just entered the service and be able to identify with some accuracy which among them will become NEAS losses at some point during their service.

5. Improvement in Data Accuracy

One of the limitations to this research is the data that was analyzed. The initial data pull generated over 500,000 observations. Once duplicate entries were dropped there were just over 300,000 observations. Upon going through the remaining data there were found to be 126,000 missing separation codes and over 16,000 missing race codes. This left just over 167,000 observations to be analyzed. This inaccuracy of data may lead to a misrepresentation of the population. The separation code is the most important variable in the study since it acts as the dependent variable. It is recommended that upon a Marine's departure from the Marine Corps an audit of records be done on that individual to ensure accurate information is present in the Total Force Data Warehouse.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX: PREDICTION MODEL RESULTS

This is an exact printout from STATA; no rounding of numbers has been applied .

****FY98-04 logit model****

Stata Command:

. logit neas_loss afqt5 afqt4 afqt3a afqt2 afqt1 nodep onedep twodep threedep fourdep fivedep six_moredep female mcd_1 mcd_4 mcd_6 mcd_8 mcd_9 mcd_12 gradeschool_educ midschool_educ jrhigh_educ college_educ master_educ postmaster_educ doctorate_educ married legal_separated married_other contract_5yr contract_6yr contract_3yr contract_8yr adult_diploma occup_cert less_highsch ged home_sch college_degree sem_college other_school combat_tour amerindian asian_pacislndr africanamerican otherrace hispanic years_0_4 years_4_8 years_8_12 age_18_19 age_20 if ecc<=16344

Results:

Logistic regression Number of obs = 105001

LR chi2(48) = 34971.76

Prob > chi2 = 0.0000

Log likelihood = -52023.019 Pseudo R2 = 0.2516

neas_loss	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
afqt5	2.145811	1.321634	1.62	0.104	-.4445427	4.736166
afqt4	.5052231	.0912804	5.53	0.000	.3263168	.6841293
afqt3a	-.1289852	.0194589	-6.63	0.000	-.1671239	-.0908466
afqt2	-.2984345	.0192431	-15.51	0.000	-.3361502	-.2607188
afqt1	-.4210761	.0446742	-9.43	0.000	-.5086359	-.3335163
nodep	-.8291134	.0217542	-38.11	0.000	-.8717509	-.7864759
onedep	.6332807	.0468561	13.52	0.000	.5414445	.7251169
twodep	-1.763581	.0310728	-56.76	0.000	-1.824483	-1.70268
threedep	-1.160006	.0449486	-25.81	0.000	-1.248103	-1.071908
fourdep	-3.075161	.0586847	-52.40	0.000	-3.190181	-2.960141
fivedep	-2.607594	.1400981	-18.61	0.000	-2.882182	-2.333007
six_moredep	-1.507096	.1264144	-11.92	0.000	-1.754864	-1.259328
female	.2378438	.0289998	8.20	0.000	.1810052	.2946823
mcd_1	.3546746	.0923981	3.84	0.000	.1735777	.5357716
mcd_4	.4861277	.0925961	5.25	0.000	.3046427	.6676128
mcd_6	.4410648	.0924082	4.77	0.000	.2599481	.6221816

neas_loss	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mcd_8	.3348314	.0923939	3.62	0.000	.1537426	.5159201
mcd_9	.3156796	.0922357	3.42	0.001	.1349009	.4964583
mcd_12	.2429449	.0924925	2.63	0.009	.0616629	.4242268
jrhigh_educ	1.927614	1.072512	1.80	0.072	-.1744708	4.029699
college_educ	-.4266084	.0469534	-9.09	0.000	-.5186354	-.3345814
master_educ	-.3641955	.3657058	-1.00	0.319	-1.080966	.3525748
postmaster~c	.2417967	1.222394	0.20	0.843	-2.154051	2.637644
married	-.5174869	.0233657	-22.15	0.000	-.5632828	-.471691
legal_sepa~d	-.3727926	.5300882	-0.70	0.482	-1.411746	.6661613
married_ot~r	.4758065	.0179388	26.52	0.000	.4406471	.510966
contract_5yr	1.381627	.0309463	44.65	0.000	1.320973	1.44228
contract_6yr	.6043978	.0548353	11.02	0.000	.4969226	.7118729
contract_3yr	.5531592	.060396	9.16	0.000	.4347852	.6715332
adult_dipl~a	.37405	.0711978	5.25	0.000	.2345049	.5135952
occup_cert	1.008208	.2576784	3.91	0.000	.5031677	1.513249
less_highsch	.4609496	.297105	1.55	0.121	-.1213656	1.043265
ged	.7992128	.0487301	16.40	0.000	.7037035	.8947221
home_sch	1.399455	.2642474	5.30	0.000	.8815399	1.917371
college_de~e	.2993909	.0655944	4.56	0.000	.1708282	.4279535
sem_college	-.1384153	.0936192	-1.48	0.139	-.3219057	.0450751
other_school	.3479638	.0778964	4.47	0.000	.1952897	.500638
combat_tour	-1.057272	.0356655	-29.64	0.000	-1.127175	-.9873688
amerindian	.0804033	.0753654	1.07	0.286	-.0673101	.2281167
asian_paci~r	-.3708884	.0625207	-5.93	0.000	-.4934267	-.2483501
africaname~n	-.0060925	.0230026	-0.26	0.791	-.0511768	.0389919
otherrace	-.3309024	.0274942	-12.04	0.000	-.3847901	-.2770147
hispanic	-.1448604	.2395963	-0.60	0.545	-.6144606	.3247398
years_0_4	-3.263197	.0491754	-66.36	0.000	-3.359579	-3.166815
years_4_8	-4.926311	.0528698	-93.18	0.000	-5.029934	-4.822688
years_8_12	-3.857037	.0572108	-67.42	0.000	-3.969168	-3.744906
age_18_19	2.433482	.8633128	2.82	0.005	.7414195	4.125544
age_20	2.758538	.8634648	3.19	0.001	1.066178	4.450898

```
   _cons        .478128     .8673074      0.55        0.581          -1.221763    2.178019
------------------------------------------------------------------------------------------------------------------
```

****FY05 loss****

```
. predict pred05 if ecc>=16355 & ecc <=16709
(option p assumed; Pr(neas_loss))
(151415 missing values generated)
```

****FY05 ROC curve****

```
. lroc if ecc>=16355 & ecc <=16709
Logistic model for neas_loss
number of observations =    15854
area under ROC curve   =   0.8591
```

****FY98-04 estat classification****

```
. estat clas
Logistic model for neas_loss

             -------- True --------
Classified |        D          ~D  |      Total
-----------+-------------------------+--------------------
    +    |     24117         9668  |     33785
    -    |    15345        55871  |     71216
-----------+-------------------------+---------------------
  Total   |   39462       65539  |     105001

Classified + if predicted Pr(D) >= .5
True D defined as neas_loss != 0
--------------------------------------------------------------
Sensitivity                  Pr( +| D)           61.11%
Specificity                  Pr( -|~D)           85.25%
Positive predictive value       Pr( D| +)   71.38%
Negative predictive value       Pr(~D| -)  78.45%
---------------------------------------------------------------
False + rate for true ~D       Pr( +|~D)    14.75%
```

False - rate for true D	Pr(-| D)	38.89%
False + rate for classified +	Pr(~D| +)	28.62%
False - rate for classified -	Pr(D| -)	21.55%

Correctly classified	76.18%

****FY98-05 logit model****

STATA Command:

. logit neas_loss afqt5 afqt4 afqt3a afqt2 afqt1 nodep onedep twodep threedep fourdep fivedep six_moredep female mcd_1 mcd_4 mcd_6 mcd_8 mcd_9 mcd_12 gradeschool_educ midschool_educ jrhigh_educ college_educ master_educ postmaster_educ doctorate_educ married legal_separated married_other contract_5yr contract_6yr contract_3yr contract_8yr adult_diploma occup_cert less_highsch ged home_sch college_degree sem_college other_school combat_tour amerindian asian_pacislndr africanamerican otherrace hispanic years_0_4 years_4_8 years_8_12 age_18_19 age_20 if ecc<=16709

Results:

Logistic regression — Number of obs = 121385

LR chi2(48) = 41219.40

Prob > chi2 = 0.0000

Log likelihood = -60181.088 — Pseudo R2 = 0.2551

neas_loss	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
afqt5	2.088865	1.303147	1.60	0.109	-.4652561	4.642986
afqt4	.4984786	.0829761	6.01	0.000	.3358484	.6611088
afqt3a	-.1420519	.0180621	-7.86	0.000	-.1774529	-.1066509
afqt2	-.303397	.0178071	-17.04	0.000	-.3382983	-.2684957
afqt1	-.4219057	.0411069	-10.26	0.000	-.5024738	-.3413375
nodep	-.8728133	.0202944	-43.01	0.000	-.9125896	-.833037
onedep	.5137931	.042678	12.04	0.000	.4301457	.5974404
twodep	-1.689209	.0277028	-60.98	0.000	-1.743505	-1.634912
threedep	-1.228228	.043488	-28.24	0.000	-1.313463	-1.142993
fourdep	-2.977514	.0514013	-57.93	0.000	-3.078258	-2.876769
fivedep	-2.66154	.1367415	-19.46	0.000	-2.929548	-2.393531
six_moredep	-1.476605	.1126954	-13.10	0.000	-1.697484	-1.255726
female	.2507925	.0266031	9.43	0.000	.1986515	.3029335
mcd_1	.4115153	.0899638	4.57	0.000	.2351894	.5878411
mcd_4	.5624521	.0901512	6.24	0.000	.3857591	.7391452
mcd_6	.4970985	.0899807	5.52	0.000	.3207396	.6734574
mcd_8	.3894488	.089986	4.33	0.000	.2130795	.5658182
mcd_9	.3739671	.0898392	4.16	0.000	.1978856	.5500486
mcd_12	.2945765	.0900601	3.27	0.001	.118062	.471091
jrhigh_educ	1.825779	.8430681	2.17	0.030	.1733958	3.478162

neas_loss	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
college_educ	-.4478418	.0441245	-10.15	0.000	-.5343242	-.3613594
master_educ	-.2880699	.359026	-0.80	0.422	-.991748	.4156082
postmaster~c	.2574373	1.215278	0.21	0.832	-2.124463	2.639338
married	-.5639686	.0222565	-25.34	0.000	-.6075906	-.5203466
legal_sepa~d	-.5030741	.5255833	-0.96	0.338	-1.533199	.5270503
married_ot~r	.3423359	.017026	20.11	0.000	.3089656	.3757062
contract_5yr	1.335812	.02822	47.34	0.000	1.280501	1.391122
contract_6yr	.4719904	.0528479	8.93	0.000	.3684105	.5755704
contract_3yr	.5134517	.0574264	8.94	0.000	.400898	.6260053
adult_dipl~a	.3839804	.066056	5.81	0.000	.2545129	.5134478
occup_cert	1.09648	.2516563	4.36	0.000	.6032427	1.589717
less_highsch	.3339898	.2953194	1.13	0.258	-.2448256	.9128051
ged	.7655964	.0453362	16.89	0.000	.6767391	.8544537
home_sch	1.661511	.2394215	6.94	0.000	1.192253	2.130768
college_de~e	.3141491	.0610762	5.14	0.000	.1944418	.4338563
sem_college	-.2434483	.0927135	-2.63	0.009	-.4251634	-.0617331
other_school	.3160142	.0737446	4.29	0.000	.1714773	.460551
combat_tour	- 1.096738	.0283235	-38.72	0.000	-1.152251	-1.041225
amerindian	.0673721	.0694278	0.97	0.332	-.0687039	.203448
asian_paci~r	-.3893545	.0576344	-6.76	0.000	-.5023158	-.2763933
africaname~n	-.0392575	.0215521	-1.82	0.069	-.0814988	.0029839
otherrace	-.2749588	.0251801	-10.92	0.000	-.324311	-.2256067
hispanic	.2151268	.1992131	1.08	0.280	-.1753237	.6055772
years_0_4	-3.299742	.045302	-72.84	0.000	-3.388532	-3.210952
years_4_8	-4.957242	.0487514	-101.68	0.000	-5.052793	-4.861691
years_8_12	-3.99003	.0531981	-75.00	0.000	-4.094296	-3.885763
age_18_19	2.478938	.8561285	2.90	0.004	.8009573	4.15692
age_20	2.814834	.8562621	3.29	0.001	1.136591	4.493077
_cons	.5558471	.8600256	0.65	0.518	-1.129772	2.241466

.

FY06 loss

. predict pred06 if ecc>=16710 & ecc <=17074

(option p assumed; Pr(neas_loss))

(150201 missing values generated)

FY06 ROC Curve

. lroc if ecc>=16710 & ecc <=17074

Logistic model for neas_loss

number of observations = 17068

area under ROC curve = 0.8773

FY98-05 classification

. estat clas

Logistic model for neas_loss

```
          -------- True --------
Classified |      D          ~D  |      Total
-----------+--------------------------+-----------
    +    |   30782      12967  |     43749
    -    |   15724      61912  |     77636
-----------+--------------------------+-----------
  Total  |  46506      74879  |     121385
```

Classified + if predicted Pr(D) >= .5

True D defined as neas_loss != 0

```
--------------------------------------------------------------
Sensitivity                   Pr( +| D)     66.19%
Specificity                   Pr( -|~D)     82.68%
Positive predictive value     Pr( D| +)     70.36%
Negative predictive value     Pr(~D| -)     79.75%
--------------------------------------------------------------
False + rate for true ~D      Pr( +|~D)     17.32%
False - rate for true D       Pr( -| D)     33.81%
False + rate for classified + Pr(~D| +)     29.64%
False - rate for classified - Pr( D| -)     20.25%
```

Correctly classified	76.36%

.

****FY98-06 logit model****

STATA Command:

. logit neas_loss afqt5 afqt4 afqt3a afqt2 afqt1 nodep onedep twodep threedep fourdep fivedep six_moredep female mcd_1 mcd_4 mcd_6 mcd_8 mcd_9 mcd_12 gradeschool_educ midschool_educ jrhigh_educ college_educ master_educ postmaster_educ doctorate_educ married legal_separated married_other contract_5yr contract_6yr contract_3yr contract_8yr adult_diploma occup_cert less_highsch ged home_sch college_degree sem_college other_school combat_tour amerindian asian_pacislndr africanamerican otherrace hispanic years_0_4 years_4_8 years_8_12 age_18_19 age_20 if ecc<=17074

Results:

Logistic regression	Number of obs	=	138455
	LR chi2(49)	=	48200.82
	Prob > chi2	=	0.0000
Log likelihood = -68304.946	Pseudo R2	=	0.2608

neas_loss	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
afqt5	1.414497	.9290875	1.52	0.128	-.4064808	3.235475
afqt4	.4916211	.0775224	6.34	0.000	.3396799	.6435622
afqt3a	-.1496628	.0169926	-8.81	0.000	-.1829678	-.1163579
afqt2	-.3009868	.0166754	-18.05	0.000	-.3336701	-.2683035
afqt1	-.4140289	.0381357	-10.86	0.000	-.4887735	-.3392843
nodep	-.9089505	.0189664	-47.92	0.000	-.9461241	-.871777
onedep	.443191	.0406111	10.91	0.000	.3635947	.5227873
twodep	-1.654814	.0250982	-65.93	0.000	-1.704005	-1.605622
threedep	-1.293341	.0425976	-30.36	0.000	-1.37683	-1.209851
fourdep	-2.854894	.0441873	-64.61	0.000	-2.941499	-2.768288
fivedep	-2.723148	.1333334	-20.42	0.000	-2.984477	-2.461819
six_moredep	-1.760051	.0950631	-18.51	0.000	-1.946372	-1.573731
female	.2232618	.0248763	8.97	0.000	.1745052	.2720185
mcd_1	.4665062	.0882612	5.29	0.000	.2935174	.639495
mcd_4	.6384017	.0884269	7.22	0.000	.4650881	.8117153
mcd_6	.5786312	.0882732	6.56	0.000	.4056188	.7516436
mcd_8	.4634792	.0882712	5.25	0.000	.2904708	.6364877
mcd_9	.4436044	.0881444	5.03	0.000	.2708446	.6163642
mcd_12	.3565509	.0883317	4.04	0.000	.183424	.5296777

neas_loss	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
midschool_~c	.8355599	2.449049	0.34	0.733	-3.964488	5.635608
jrhigh_educ	2.264275	.6963263	3.25	0.001	.8995007	3.62905
college_educ	-.426604	.0414272	-10.30	0.000	-.5077999	-.3454081
master_educ	-.2636766	.3525344	-0.75	0.454	-.9546314	.4272782
postmaster~c	.9198944	1.116911	0.82	0.410	-1.269211	3.108999
married	-.5910637	.0212955	-27.76	0.000	-.6328022	-.5493252
legal_sepa~d	-.4789016	.4951528	-0.97	0.333	-1.449383	.49158
married_ot~r	.2329338	.0163924	14.21	0.000	.2008054	.2650623
contract_5yr	1.300564	.0259389	50.14	0.000	1.249724	1.351403
contract_6yr	.4250908	.0507025	8.38	0.000	.3257158	.5244657
contract_3yr	.4645321	.0537219	8.65	0.000	.3592391	.5698251
adult_dipl~a	.4112988	.061882	6.65	0.000	.2900123	.5325854
occup_cert	1.107374	.2471434	4.48	0.000	.6229816	1.591766
less_highsch	.2378835	.2941645	0.81	0.419	-.3386683	.8144353
ged	.7533898	.0432021	17.44	0.000	.6687152	.8380645
home_sch	1.588673	.1946711	8.16	0.000	1.207124	1.970221
college_de~e	.2970497	.0568641	5.22	0.000	.1855981	.4085012
sem_college	-.324012	.0920348	-3.52	0.000	-.5043969	-.1436272
other_school	.2853973	.0710383	4.02	0.000	.1461648	.4246299
combat_tour	-1.024912	.0237086	-43.23	0.000	-1.07138	-.978444
amerindian	.0395909	.06484	0.61	0.541	-.0874932	.1666751
asian_paci~r	-.3825699	.053652	-7.13	0.000	-.4877258	-.277414
africaname~n	-.0609111	.0204943	-2.97	0.003	-.1010793	-.0207429
otherrace	-.2274422	.0234664	-9.69	0.000	-.2734355	-.1814488
hispanic	.2339587	.1628157	1.44	0.151	-.0851542	.5530716
years_0_4	-3.267119	.0420151	-77.76	0.000	-3.349467	-3.184771
years_4_8	-4.922813	.045257	-108.77	0.000	-5.011515	-4.834111
years_8_12	-4.085572	.0495954	-82.38	0.000	-4.182777	-3.988367
age_18_19	2.479552	.8446311	2.94	0.003	.8241056	4.134999
age_20	2.820221	.8447512	3.34	0.001	1.164539	4.475903
_cons	.5727886	.848517	0.68	0.500	-1.090274	2.235851

.

****FY07 loss****

```
. predict pred07 if ecc>=17075 & ecc <=17439
(option p assumed; Pr(neas_loss))
(151464 missing values generated)
```

****FY07 ROC curve****

```
. lroc if ecc>=17075 & ecc <=17439
Logistic model for neas_loss
number of observations =    15805
area under ROC curve   =   0.8837
```

****FY98-06 classification****

```
. estat clas
Logistic model for neas_loss
           -------- True --------
Classified |        D          ~D  |      Total
-----------+-------------------------+----------------
    +    |   37437       15912  |      53349
    -    |   16150       68956  |      85106
-----------+-------------------------+----------------
  Total   |  53587       84868  |      138455

Classified + if predicted Pr(D) >= .5
True D defined as neas_loss != 0
--------------------------------------------------------------
Sensitivity                   Pr( +| D)           69.86%
Specificity                   Pr( -|~D)           81.25%
Positive predictive value       Pr( D| +)   70.17%
Negative predictive value       Pr(~D| -)  81.02%
----------------------------------------------------------------
False + rate for true ~D        Pr( +|~D)    18.75%
False - rate for true D         Pr( -| D)       30.14%
False + rate for classified +   Pr(~D| +)   29.83%
False - rate for classified -   Pr( D| -)       18.98%
----------------------------------------------------------------
Correctly classified                           76.84%
```

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Hattiangadi, A.U., Kimble, T.H., Lambert, W.B., Quester, A.O. (2005). Center for Naval Analysis. *Endstrength: Forecasting Marine Corps Losses Final Report*. Alexandria,VA: CNA.

Hawes, E.A., (1990). *An Application of Survival Analysis Methods to the Study of Marine Corps Enlisted Attrition*. United States Naval Postgraduate School.

Hurst, S.F., Manion, T.E., (1985). *Use of Binary Choice Model to Determine Marine Officer Attrition*. United States Naval Postgraduate School.

Klein, S., Hawes-Dawson, J., Martin, T., (1991). *Why Recruits Separate Early*. Santa Monica, CA: RAND.

Total Force Data Warehouse. https://tfdw-web.manpower.usmc.mil. Accessed: November 2007.

Wooldridge, J.M., Introductory Econometrics, A Modern Approach. Third Edition. Thomson South Western. Mason, OH. (2006).

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Marine Corps Representative
   Naval Postgraduate School
   Monterey, California

4. Samuel Buttrey
   Naval Postgraduate School
   Monterey, California

5. Stephen Mehay
   Naval Postgraduate School
   Monterey, California

6. Wayne Wagner
   Strategic Affairs Office, N1Z
   Arlington Annex, Washington District of Columbia

7. Sanford Orrick
   Fredericksburg, Virginia